**Protocol for a Systematic Literature Review on the topic of High-value datasets**

The protocol is intended for the Systematic Literature review on the topic of High-value Datasets with the aim to gather information on how the topic of High-value datasets (HVD) and their determination has been reflected in the literature over the years and what has been found by these studies to date, incl. the indicators used in them, involved stakeholders, data-related aspects, and frameworks.

To achieve the research objective, the following research questions (RQ) were established:

* *(RQ1) how is the value of the open government data perceived / defined? In which contexts has the topic of HVD been investigated by previous research (e.g., research disciplines, countries)? Are local efforts being made at the country levels to identify the datasets that provide the most value to stakeholders of the local open data ecosystem?*
* *(RQ1.1) how the high-value data are defined, if this definition differs from the definition introduced in the PSI /OD Directive, and (RQ1.2) what datasets are considered to be of higher value in terms of data nature, data type, data format, data dynamism?*
* *(RQ2) What indicators are used to determine high-value datasets? How can these indicators be classified? Can they be measured? And whether this can be done (semi-)automatically?*
* *(RQ3) Whether there is a framework for determining country specific HVD? In other words, is it possible to determine what datasets are of particular value and interest for their further reuse and value creation, taking into account the specificities of the country under consideration, e.g., culture, geography, ethnicity, likelihood of crises and/or catastrophes.*

To attain the objective of our study, we developed the protocol[[1]](#footnote-1), where the information on each selected study was collected in four categories: (1) descriptive information, (2) approach- and research design- related information, (3) quality-related information, (4) HVD determination-related information (see Table below).

|  |  |  |
| --- | --- | --- |
| **Category** | **Metadata** | **Description** |
| **Descriptive information** | Article number | A study number, corresponding to the study number assigned in an Excel worksheet |
| Complete reference | The complete source information to refer to the study |
| Year of publication | The year in which the study was published |
| Journal article / conference paper / book chapter | The type of the paper, i.e., journal article, conference paper, or book chapter |
| DOI / Website | A link to the website where the study can be found |
| Number of citations | The number of citations of the article in Google Scholar, Scopus, Web of Science |
| Availability in OA | Availability of an article in the Open Access |
| Keywords | Keywords of the paper as indicated by the authors |
| Relevance for this study | What is the relevance level of the article for this study? (high / medium / low) |
| **Approach- and research design-related information** | Objective / RQ | The research objective / aim, established research questions |
| Research method (including unit of analysis) | The methods used to collect data, including the unit of analysis (country, organisation, specific unit that has been analysed, e.g., the number of use-cases, scope of the SLR etc.) |
| Contributions | The contributions of the study |
| Method | Whether the study uses a qualitative, quantitative, or mixed methods approach? |
| Availability of the underlying research data | Whether there is a reference to the publicly available underlying research data e.g., transcriptions of interviews, collected data, or explanation why these data are not shared? |
| Period under investigation | Period (or moment) in which the study was conducted |
| Use of theory / theoretical concepts / approaches | Does the study mention any theory / theoretical concepts / approaches? If any theory is mentioned, how is theory used in the study? |
| **Quality- and relevance- related information** | Quality concerns | Whether there are any quality concerns (e.g., limited information about the research methods used)? |
| Primary research object | Is the HVD a primary research object in the study? (primary - the paper is focused around the HVD determination, secondary - mentioned but not studied (e.g., as part of discussion, future work etc.)) |
| **HVD determination-related information** | HVD definition and type of value | How is the HVD defined in the article and / or any other equivalent term? |
| HVD indicators | What are the indicators to identify HVD? How were they identified? (components & relationships, “input -> output") |
| A framework for HVD determination | Is there a framework presented for HVD identification? What components does it consist of and what are the relationships between these components? (detailed description) |
| Stakeholders and their roles | What stakeholders or actors does HVD determination involve? What are their roles? |
| Data | What data do HVD cover? |
| Level (if relevant) | What is the level of the HVD determination covered in the article? (e.g., city, regional, national, international) |

1. The protocol structure is based on Zuiderwijk, A., Chen, Y. C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. Government Information Quarterly, 38(3), 101577. [↑](#footnote-ref-1)